图书情报工作
第 64 卷 第 22 期　2020 年 11 月

科学数据开放共享中的数据安全治理研究*

■ 盛小平　郭道胜
上海大学图书情报档案系　上海 200444

摘　要：[目的/意义]揭示科学数据开放共享中的数据安全问题，提出相应的治理对策，以更好地促进我国科学数据开放共享实践。[方法/过程]运用规范分析法，梳理与界定科学数据开放共享中的数据安全问题，然后从机密性、完整性、可用性 3 个维度探究科学数据安全治理措施。[结果/结论]科学数据开放共享在数据机密性、完整性和可用性方面存在许多安全问题。加强数据安全立法、建立科学数据分级分类标准与系统、充分利用隐私增强技术 3 项措施可以治理数据机密性问题；建立数据保护官制度、实施数据保护影响评估、运用数据认证技术 3 项措施可以治理数据完整性问题；制定科学数据可用性政策、提高科学数据质量、构建基于数据联盟的国家科学数据中心 3 项措施可以治理数据可用性问题。

关键词：科学数据　开放共享　数据安全　安全治理
分类号：G203
DOI：10.13266/j.issn.0252-3116.2020.22.003

1　引言

大数据时代，数据作为支撑与推动各国产业与创新发展的核心资产，受到前所未有的重视与保护。数据开放共享与数据安全治理成为“一个硬币的两面”，也是各国政策法律关注的焦点[1]。被誉为“史上最严格”数据保护法案的欧盟《一般数据保护条例》(General Data Protection Regulation，GDPR)于 2018 年 5 月 25 日正式实施，成为全球数据安全保护的重要标杆。在 GDPR 基础上，英国颁布了《2018 数据保护法》(Data Protection Act 2018)；我国已经颁布实施了《中华人民共和国国家安全法》(以下简称为《国家安全法》)、《中华人民共和国网络安全法》(以下简称为《网络安全法》)，《个人信息保护法》《中华人民共和国数据安全法》(以下简称《数据安全法》)也被纳入十三届全国人大常委会的立法规划，《数据安全管理办法》已于 2019 年 6 月 28 日完成了公开征求意见。数据安全被快速提升到重要高度，每个人、每个企业、每个行业都无法置身事外[2]。近几年来，人们已经对科学数据开放共享[3-4]、开放数据保护[5]、数字数据保护[6]、数据安全管理[7]、数据安全治理[8-9]、数据保护与治理[10]、数据安全风险治理[11]、数据隐私管理[12]、大数据隐私与安全政策[13]等主题进行了广泛探索。然而，鲜见成果深入论述科学数据开放共享中的数据安全治理问题。事实上，数据安全治理与数据安全管理是两个不同概念。数据安全管理是对安全策略和程序的规划、开发和执行，以提供数据和信息资产的适当认证、授权、访问和审计。数据安全管理的基本目标是要确保合适的人以正确的方式使用和更新数据，并限制所有不适当的访问和更新数据；数据安全管理的最终目标是保护数据资产符合隐私与保密法规要求，并与业务要求相一致[14]。数据安全治理是维护组织数据资产的机密性、完整性和可用性的系统，包括管理承诺和领导、组织结构、用户意识和承诺、政策、程序、流程、技术和合规执行机制[15]；也是对数据安全进行综合治理的过程，它需要从决策层到技术层、从管理制度到工具支撑，自上而下在各个层级之间对数据安全治理的目标达成共识，确保采取合理和适当的措施，以最有效的方式保护数据资产[16]。数据安全治理的主要目标是确保组织数据资产的安全性，并实现数据资产的保值与增值。数据安全管理的主要业务活动是理解组织数据需求和监管要求；定义数据安全策略和标准；定义数据安全控

* 本文系国家社会科学基金项目“开放科学环境下的科学数据开放共享机制与对策研究”(项目编号：18ATQ007)研究成果之一。
作者简介：盛小平(ORCID：0000-0002-6341-6973)，教授，博士，博士生导师，E-mail：shengxp68@126.com；郭道胜(ORCID：0000-0001-7181-9484)，硕士研究生。
收稿日期：2020-06-09　修回日期：2020-08-17　本文起止页码：25-36　本文责任编辑：王传清

制及措施；管理用户和密码及访问权限；监控用户身份认证和访问行为；划分数据与信息等级；审计数据安全[14]152-160。而数据安全治理的主要业务活动是理解组织数据安全战略需求；发展和维护组织数据安全战略；建立数据安全治理机构与制度；任命数据安全管理专员；制定并审核数据安全政策、标准和程序；协调数据安全治理活动；解决数据安全相关问题；监督数据安全管理项目与服务；评估数据资产价值；监控合规行为。由此看来，尽管数据安全管理与数据安全治理有内在联系，但两者在主要目标、业务活动等方面存在明显差异。总之，数据安全管理为数据安全治理奠定基础，数据安全治理为数据安全管理提供保障。由于解决数据安全问题对于我国实施科学数据开放共享至关重要，在目前对科学数据开放共享中的数据安全治理问题缺少相关研究的情况下，本文将在界定科学数据开放共享中的数据安全问题基础上，构建科学数据开放共享中的数据安全治理模型与对策，以期更好地促进科学数据开放共享和开放创新。

2 科学数据开放共享中数据安全问题的界定

欧洲委员会要求，开放研究数据试验项目必须保存那些支持发表在同行评审出版物上的研究结果的数据和他们定义的其他数据，最好是存入研究数据存储库，并采取措施使这些研究数据能被第三方访问、挖掘、利用、重新制作和被任何用户免费传播[17]。2018年3月17日，国务院办公厅颁布了《科学数据管理办法》，以促进科学数据的开放共享。那么，在科学数据开放共享过程（即开放获取、开放存储、开放发布、开放利用）中，数据安全就成为一个不可忽视的问题。这首先需要对数据安全概念和科学数据开放共享存在哪些主要的数据安全问题有深入的认识。

2.1 数据安全概念与内涵

数据安全是一门研究如何保护计算机和通信系统中的数据免受未经授权的泄露和修改的科学，包含密码控制、访问控制、信息流控制、推理控制等4种控制活动以及备份和恢复过程[18]。数据安全可分为物理、人员、程序与技术4个维度，见表1。

经典的数据安全需求是数据机密性、完整性和可用性等，其目的是防止数据在传输、存储等环节中被泄露或破坏[20]。数据机密性意味着一个安全系统仅允许个人看到其可以看到的数据，包括保证数据通信的

表1 数据安全维度[19]

维度	安全问题
物理	未经授权的用户必须在物理上无法访问你的计算机。这意味着你必须将数据保存在安全的物理环境中
人员	负责系统管理和数据安全的人员必须是可靠的，且在雇用数据库管理员前需要检查其背景
程序	系统运行中使用的程序能够确保可靠的数据
技术	数据的存储、访问、操作和传输必须受到技术保护，这些技术可以增强特定信息控制策略

隐私、实现敏感数据的安全存储、能够验证有效的用户和实施粒度访问控制。数据完整性是指数据的一致性、正确性、有效性和相容性，意味着数据存储在数据库中或通过网络传输数据时，能够得到保护而不被删除和损坏。数据可用性意味着一个安全系统授权用户可以不受延迟地访问数据。由于数据或信息是现代组织的核心资产，其机密性、完整性和可用性是21世纪任何组织长期生存的基础，所以任何组织除非采取全面和系统的办法来保护其数据或信息的机密性、完整性和可用性，否则它们将容易受到各种可能的威胁[21]。这包括威胁数据安全的多种情形，比如硬盘驱动器损坏、人为错误或操作失误、黑客入侵、病毒感染、信息窃取、自然灾害、电源故障、磁干扰等。不过，数据安全需要澄清和纠正过去一些安全神话[19]：①黑客造成了大多数安全漏洞。事实上，80%的数据损失是由内部人士造成的。②加密使你的数据安全。事实上，加密只是保护数据的一种方法。安全性还需要访问控制、数据完整性、系统可用性和审核。③防火墙使你的数据安全。事实上，40%的互联网入侵者都是在设置了防火墙的情况下发生的。

2.2 科学数据开放共享中的数据安全问题

科学数据开放共享中的数据安全问题同样体现在数据机密性、完整性和可用性3方面。

2.2.1 有关科学数据机密性的安全问题

科学数据开放共享需要保证科学数据的机密性。目前在科学数据开放共享中，涉及机密性的数据安全问题主要包括：①隐私泄露，缺少有效的隐私保护，比如公共卫生领域的许多研究涉及医疗记录和病史，这使得在开放共享研究成果的同时保护患者的隐私变得非常困难。②匿名数据并不十分安全。开放共享将使数据控制者失去对谁可以访问数据的控制。即使是匿名的数据，也可以显示出有关数据主体的私人信息，或可能仍然包含与个人有关的敏感信息，通过将这些数据与其他可公开获得的信息联系起来，可以重新确定个人的身份[22]，因此匿名数据并不完全安全[5]25-26。

盛小平，郭道胜. 科学数据开放共享中的数据安全治理研究[J]. 图书情报工作,2020,64(22):25－36.

③科学数据开放共享与个人数据保护存在某种程度的冲突。开放共享需要把包含个人信息的科学数据存储在开放数据知识库中,能让用户不受任何限制地获取、挖掘、复制、传播和利用。这显然与个人数据保护原则相冲突[5]193。④缺少数据分类分级规范和标准。目前我国尚缺少对数据开放共享的顶层设计,尚未建立政府数据、科学数据分类分级的规范和标准,无法有效识别重要数据、敏感数据和隐私数据,缺少针对不同类型数据开放的指导原则[23]。⑤知识产权保护机制不完善[24],比如:数字科学数据知识产权难以界定[25];缺乏科学数据开放共享的法律框架;某些数据共享和数据使用权的法律条款相互矛盾[26]。⑥没有采用有效的数据安全保护技术,如没有采用数据加密或增强隐私等技术对科学数据实施有效的保护。

2.2.2 有关科学数据完整性的安全问题

科学数据开放共享需要确保科学数据的完整性,但面临如下一些关键挑战:①数据格式不标准、不一致,或数据不完整、过于复杂,数据软件不兼容。②科学数据结果冲突,如利用同一数据在相同条件下产生的研究结果相矛盾,存储在不同系统中的相似数据产生不同的结果[27]。③数据污染,如数据失真、数据造假、数据超载等[28]。④数据窃取与篡改。某些个人或机构为了商业利益或其他不良目的,可能窃取开放共享的科学数据,既不说明数据的来源,也不标注数据参考的文献,甚至使科学数据失去可靠性。⑤数据滥用。开放共享的科学数据为不法之徒滥用数据提供了可能,比如泄露科学数据中涉及的敏感个人信息、商业机密或国家情报来换取商业报酬。⑥数据丢失,比如:部分辅助数据缺失或不全面,历史数据丢失严重[29];因研究笔记本被丢弃造成的原始数据遗失;计算机硬盘崩溃造成的数据损坏;数字媒介随着时间的推移出现的衰退等[30]。

2.2.3 有关科学数据可用性的安全问题

科学数据开放共享需要保证科学数据的可用性,其安全问题主要包括:①科学数据没有得到妥善记录和处理。许多科学数据集可能从一开始就没有被记录和存储下来,从而不能再使用。芬兰社会科学数据档案馆调查发现,54% 受访者认为对数据可用性(如不完整的文档)的担心是使得数据在其领域没有再利用的重要原因[31]。②个人数据用于科学研究受到较强限制。科学研究豁免不能使个人数据的处理合法化,只能使较长的存储周期合法化或进一步的处理合法化[5]201－202。研究者在处理个人数据时必须得到数据所有者的同意。③缺少完善的科学数据开放共享平台,国内科学数据共享平台建设整体情况不优,网站功能较为单一,可浏览、检索、获取的数据资源较少[32]。④数据间彼此孤立,数据更新没有保障,数据可用性较差。⑤数据权利模糊,缺少有效授权。有关科学数据的知情权、采集权、所有权、保存权、使用权等数据权益归属目前模糊不清,没有得到法律的有效界定[31],也未解决多作者数据共有权问题[33]。芬兰社会科学数据档案馆调查发现,47% 的受访者认为,缺乏所有权协议是数据不被重复使用的一个重要原因;三分之二(66%)的受访者认为,"缺乏知情同意"是开放获取研究数据的一个主要障碍;48% 的人认为开放获取增加了与机密性、研究道德和数据保护有关的风险[31]9－10,12－13。

3 科学数据开放共享中的数据安全治理模型构建

为解决上述种种科学数据开放共享中的数据安全问题,加强数据安全治理是一种应然的选择。由于上述数据安全问题既涉及科学数据开放共享的许多主要活动,比如科学数据的开放获取、开放存储、开放发布、开放利用(如开放引用)等,又涉及数据安全治理的许多关键环节,包括数据产权保护、数据隐私保护、数据安全监控、数据质量监控、数据设施建设、数据人员管理、数据安全技术开发与应用等,因此,这里借鉴迈克尔－波特(M. E. Porter)的价值链模型[34],构建了科学数据开放共享数据安全治理模型,如图 1 所示:

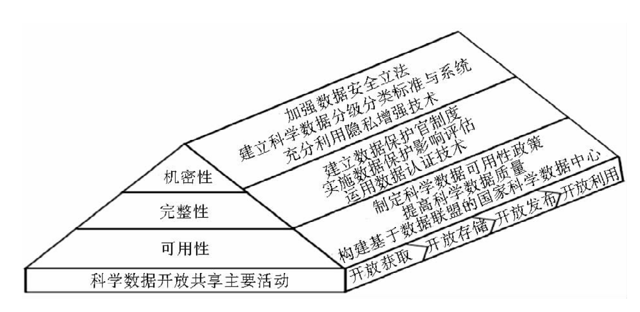


图 1　科学数据开放共享数据安全治理模型

该模型的主要特征和价值是:①贯彻了价值链思想。该模型将科学数据开放共享价值链的主要活动概括为开放获取、开放存储、开放发布和开放利用,并把科学数据开放共享中的各种数据安全治理对策作为科学数据开放共享价值链的辅助活动,从数据安全需求维度整合了这些主要活动与辅助活动,使科学数据开放共享中的数据安全问题与治理对策融为一体。②强

调了问题导向。该模型以科学数据的机密性、完整性和可用性需求为导向,构建了面向不同数据安全需求的数据安全治理对策,使科学数据开放共享中可能出现的各种数据安全问题都可找到合适的治理对策。③强调了关联互动与集成。该模型形成了一个覆盖主要科学数据开放共享活动的相互关联的数据安全治理体系,能够适用于各种科学数据开放共享环境。

4 科学数据开放共享中的数据安全治理对策

基于上述的数据安全治理模型,可以从机密性、完整性、可用性 3 个维度解析数据安全治理对策,以促进科学数据开放共享。

4.1 面向数据机密性的治理对策

由于科学数据开放共享遇到的数据机密性问题既涉及到管理机制不健全,如缺少数据保护法、缺少数据分类分级标准,也涉及到数据保护技术不到位,如存在隐私泄露、没有采用增强隐私技术等,因此,这里从立法、管理与技术 3 方面提出 3 种治理措施。

4.1.1 加强数据安全立法,夯实科学数据安全治理的法律基础

为有效保护数据安全,欧盟 2016 颁布了 GDPR,德国 2017 年通过了新版《德国联邦数据保护法》,英国 2018 年通过了新版《数据保护法》。尽管我国已制定实施了《国家安全法》《网络安全法》《科学数据管理办法》等,但迄今为止没有颁布一部专门的《数据安全法》或《数据保护法》。2015 年 7 月 1 日开始实施的《国家安全法》构建了集政治安全、国土安全、军事安全、经济安全、文化安全、社会安全、科技安全、网络与信息安全、生态安全、资源安全、太空安全、深海安全、极地安全、核安全等于一体的国家安全体系[35],但该法没有明确政府、机构或个人如何保障科学数据的安全。2017 年 6 月 1 日起施行的《网络安全法》规范了网络层面的安全要求,如明确要求需要维护网络数据的完整性、保密性和可用性;网络运营者应当保障网络免受干扰、破坏或者未经授权的访问,防止网络数据泄露或者被窃取、篡改,不得泄露、篡改、毁损其收集的个人信息,未经被收集者同意,不得向他人提供个人信息;任何个人和组织不得从事窃取网络数据等危害网络安全的活动,不得提供专门用于从事侵入网络、干扰网络正常功能及防护措施、窃取网络数据等危害网络安全活动的程序、工具[36]等,但它难以系统性解决数据安全保障问题[37]。2018 年 3 月 17 日颁布实施的《科学数据管理办法》明确规定:涉及国家秘密、国家安全、社会公共利益、商业秘密和个人隐私的科学数据,不得对外开放共享;确需对外开放的,要对利用目的、用户资质、保密条件等进行审查,并严格控制知悉范围;主管部门和法人单位应建立健全涉及国家秘密的科学数据管理与使用制度,对制作、审核、登记、拷贝、传输、销毁等环节进行严格管理;主管部门和法人单位应加强科学数据全生命周期安全管理,制定科学数据安全保护措施,加强数据下载的认证、授权等防护管理,防止数据被恶意使用;主管部门和法人单位对于需对外公布或提供的科学数据应建立相应的安全保密审查制度;法人单位和科学数据中心应建立网络安全保障体系,健全防篡改、防泄露、防攻击、防病毒等安全防护体系;科学数据中心应建立应急管理和容灾备份机制以及应急管理系统,对重要的科学数据进行异地备份[38]。然而,《科学数据管理办法》并没有明确在不同科学数据生命周期阶段如何实施有效的数据安全管理,更没有针对科学数据开放共享行为规划数据安全治理措施。因此,科学数据开放共享中的数据安全问题仍缺少充足的法律保障。

这种现象已经引起了立法机构、政府部门和专家的重视。2019 年 5 月 28 日,“国家互联网信息办公室关于《数据安全管理办法(征求意见稿)》公开征求意见的通知”对外公开发布。该征求意见稿主要规范了网络运营者在境内利用网络收集、存储、传输、处理、使用数据的行为以及数据安全监管要求[39],有助于强化和明确网络运营者在保障数据安全方面的责任,并有效发挥其作用。但是,它仍无法全面解决科学数据开放共享中的数据安全问题,主要缺陷之一是它没有规范其他利益相关者(如数据生产者、组织者、使用者等)在确保数据安全上的责任与监督保障机制。可喜的是,2020 年 6 月 28 – 30 日,第十三届全国人大常委会第二十次会议审议了《数据安全法(草案)》,并于 2020 年 7 月 2 日向社会公布,征求公众意见。《数据安全法(草案)》共 7 章 51 条,涉及总则、数据安全与发展、数据安全制度、数据安全保护义务、政务数据安全与开放、法律责任等内容。其中,“建立健全数据安全治理体系,提高数据安全保障能力”和“国家建立健全数据安全协同治理体系”分别被列入第四条和第九条条文中。总之,希望国家尽快出台《数据安全法》,以夯实我国科学数据安全治理的法律基础。

4.1.2 建立科学数据分级分类标准与系统，实现科学数据安全合理管控

数据分级分类在数据安全治理过程中具有重要意义。数据分类是根据数据的共同特征，例如它们的敏感性水平和风险以及保护它们的合规规则，将数据分离和组织到相关组或类的过程[40]。2015 年国务院发布的《促进大数据发展行动纲要》明确要求"推进数据采集、政府数据开放、指标口径、分类目录、交换接口、访问接口、数据质量、数据交易、技术产品、安全保密等关键共性标准的制定和实施"[41]。我国《科学数据管理办法》规定，科学数据中心需负责科学数据的分级分类、加工整理和分析挖掘；法人单位要对科学数据进行分级分类，明确科学数据的密级和保密期限、开放条件、开放对象和审核程序等，按要求公布科学数据开放目录，通过在线下载、离线共享或定制服务等方式向社会开放共享[38]。在此背景下，科学数据分级分类管理已经成为亟待解决的一个关键问题。

最近，我国工业和信息化部颁布了《工业数据分类分级指南（试行）》，率先在国内把工业数据分为一、二、三、四级，鼓励企业在做好数据管理的前提下适当共享一、二级数据，但二级数据只对确需获取该级数据的授权机构及相关人员开放，三级数据原则上不共享，确需共享的应严格控制知悉范围[42]。尽管工业数据与科学数据并非同一个概念，但工业数据作为一种重要的科学数据，该指南可以为制定科学数据分级分类标准提供参考。

不过，科学数据种类繁多，包括但不限于[43]：研究过程中产生的任何数据；对研究人员进行的研究有重要意义的任何记录数据；验证研究结果可能需要的来源或主要资料；研究过程中获得和生成的数字对象集；应用程序内容（如分析软件、模拟软件、模式的输入和输出等）；数据库内容（视频、音频、文本、图像）；研究项目的监管数据；设计作品集和实物模型；研究日志文档；实验结果与实验室笔记、现场笔记、日记内容；参考书目和阅读资料；电子表格；元数据；方法和工作流程；模型、算法、脚本；笔记、录音带、录像带；乐谱草稿；人类、动物、地质资料；图像或数据可视化；照片、电影；植物资料，细胞、细菌、病毒样本或标本；治疗的临床记录和检测结果；蛋白质或基因序列；问卷、成绩单、密码本；各种采访记录；对调查或问卷的回应；调查结果；测试反应；光谱；标准操作程序和协议；商业秘密、商业信息、在发布前需要保密的资料，或受法律保护的类似信息。因此，需要制定专门的科学数据分级分类标准。

2017 年，美国华盛顿特区采用了一种 5 级数据分类模式，即 0 级（开放数据）、1 级（公共数据）、2 级（供地方政府使用的数据）、3 级（机密数据）、4 级（限制机密数据）[44]，受到了开放数据倡导者的广泛赞扬。加州大学伯克利分校把研究数据分为：1 级（敏感性最小的，即公共信息）、2 级（低度敏感性，即非公共、非敏感的个人身份信息）、3 级（中度敏感性的个人可识别信息）、4 级（高度敏感的个人可识别信息）[45]。类似地，澳大利亚新南威尔士大学把数据分为公共级、私人级、敏感级、高度敏感级 4 个层次[46]。基于这些观点，可以把科学数据分为如下 4 级，如表 2 所示：

表 2 科学数据分级

等级代码	等级描述	数据类型样本
4 级	高度机密的科学数据	• 受法律、法规或合同保护而不能公开的科学数据 • 与国家安全有关而不能公开的科学数据 • 与国家或机构核心技术有关而不能公开的科学数据 • 一旦泄露将产生重大影响的科学数据
3 级	机密的科学数据	• 受法律、法规或合同限制不得公开的科学数据 • 与商业秘密有关的科学数据 • 与隐私保护有关的科学数据 • 一旦泄露将产生中度影响的科学数据
2 级	有限共享的科学数据	• 仅在机构范围内可共享的科学数据 • 仅在项目团队内可共享的科学数据 • 仅在"合理使用"条件下可共享的科学数据 • 仅在特殊时期内可共享的科学数据
1 级	开放共享的科学数据	• 在互联网上可以被公众免费利用的科学数据 • 作为公共产品的科学数据 • 一般公开的数据

为有效实施科学数据分级分类，可以利用"数据标签"（datatags）来构建科学数据分级分类系统。该系统实质上是利用数据标签将数据分级、数据安全属性、数据访问条件彼此关联起来，建立一个数据标签知识库以根据不同级别的安全和访问需求存储与共享数据文件和实现数据分级分类管理[47]。基于上述的科学数据分级表，可以构建一个具有 4 个级别的数据标签分级模型（见表 3）。该模型用 4 种不同颜色代表不同的标签类别，对应于不同级别的数据。随着级别的增加，传输、存储和访问需求以及安全属性也会增加。例如，在最低级别，蓝色数据标签不需要访问凭证。绿色数据标签要求验证请求者的电子邮件地址，可能需要在电子邮件消息中发送一个链接，请求者必须响应该链接；或者使用密码凭证。从黄色数据标签起，请求者必须签署数据使用协议，使用密码或身份验证。红色数据标签需要双因素授权，如可能需要同时验证请求者的电子邮件和移动电话号码。利用计算机，可以实现数据标签自动化处理，这有助于实现科学数据安全的合理管控。

表 3 科学数据标签分级模型

标签类别/数据等级	标签描述	安全属性	访问凭证
红色/4 级	高度机密的科学数据	多重加密存储,机密传输	双因素认证,签署数据使用协议
黄色/3 级	机密的科学数据	加密存储、机密传输	密码或身份验证,签署数据使用协议
绿色/2 级	有限共享的科学数据	明文存储与传输	邮箱或密码注册,点击数据使用协议
蓝色/1 级	开放共享的科学数据	明文存储与传输	无需凭证

4.1.3 充分利用隐私增强技术,加强科学数据机密性保护

科学数据的开放共享,使得科学数据更加公开透明,可以被科学研究者再次利用,对于营造良好的科学研究环境有巨大的促进作用,但也揭露了其隐匿性,为隐私数据带来了威胁。在这种情况下,潜在的侵犯数据隐私行为不可避免地增加了。这时可利用隐私增强技术来降低数据隐私风险,增强数据机密性保护。隐私增强技术有多种多样,包括差分隐私、联合分析、同态加密、零知识证明、功能加密、安全多方计算、可搜索加密、私人信息检索、智能合约等。其中差分隐私、同态加密、零知识证明和安全多方计算 4 项技术尤其受人关注。对数据进行匿名处理虽然可以保护敏感数据,但匿名化技术需要依赖于背景知识假设,这使得往往只能保证单一数据集上隐私数据不被泄露,无法满足海量数据环境下对于数据隐私的保护。差分隐私技术克服了匿名化技术必须了解目标外部信息的缺陷,可应用于科学数据共享领域中。差分隐私保护通过对真实数据添加随机扰动,使保护对象数据失真且同时保持数据集中特定数据或数据属性(如统计特性等)不变,由此保证数据在被干扰后仍具有一定的可用性而实现隐私保护的目的[48]。同态加密技术允许对加密文本执行计算,加密文本将生成一个加密结果,该结果将与使用未加密的原始数据得到的结果一样[49]。这使得一方面可以不必在传输科学数据的同时提供对应的密钥,避免了科学数据在传输过程中被拦截或窃取所造成的数据泄露;另一方面可以为科学研究过程中的实时合作提供契机,科学研究者相互之间可以共享科学数据而不必担心原始数据泄露。零知识证明与同态加密有些类似,都不泄露任何的原始数据。这项技术可以验证信息是否有效,而无需暴露证明该信息的数据[50]。这为判断科学数据的使用是否始终与申请敏感数据的初始目的保持一致提供了可能,可以防止敏感科学数据被滥用。安全多方计算是一种加密协议,可在多方之间分配计算,允许相互不信任的各方在其私有数据上进行合作计算[51]。安全多方计算技术的实现并没有任何规定,常常会用到同态加密、混淆电路、不经意传输等技术。对于科学数据开放共享而言,安全多方计算带来的最大裨益是满足并超过了 GDPR 中对于跨境数据传输的要求,因为安全多方计算可以使数据科学家和研究人员能够对分布式数据进行合规、安全和私密的计算,而无需暴露或移动它们。总之,隐私增强技术具有巨大的潜力且近年来发展迅速,可以为科学数据开放共享中的数据机密性保护提供支持。

4.2 面向数据完整性的治理对策

在科学数据开放共享活动中,常见的可能对数据完整性造成破坏的威胁主要有[52]:①硬件故障:存储设备或其他计算机硬件故障可能导致损坏。②配置问题:计算系统(例如软件或安全应用程序)中的配置错误会损坏数据。③人为错误:人们会犯错误,并可能会意外损坏数据。④传输中的损坏:数据在传输到存储设备或通过网络传输时可能会损坏。⑤故意破坏:人或软件侵入计算机并更改数据。这 5 类威胁源自数据管理不善、数据保护不周、相关技术缺失等。因此,可以通过建立数据保护官制度、实施数据保护影响评估、运用数据认证技术来加强数据完整性问题治理。

4.2.1 建立数据保护官制度

为了加强数据安全管理与保护,GDPR 明确要求进行数据处理的政府部门或公共机构、以大规模数据处理作为核心业务(包括对数据进行定期、常态、系统监测和处理)的机构、拥有 250 名或以上员工的企业设立数据保护官(data protection officer,DPO)岗位。DPO 是负责监督某个组织的数据保护战略及其实施,并确保该组织遵守数据保护法律法规要求的官员。DPO 的职责包括但不限于[53-54]:①起草、审查和更新数据保护政策。②为可能影响多个部门使用个人数据的决策提供重点,包括进行数据保护(或隐私)影响评估。③与负责组织内相关事务和职能的其他适当人员协调。④管理个人数据处理业务可能出现的任何风险,同时考虑到处理的性质、范围、背景和目的。⑤持续进行控制评估以确保遵守关键数据保护程序。⑥以适当

方式处理和管理与个人数据保护有关的查询和来自数据主体的投诉,包括本组织为处理投诉而应采取的任何行动。⑦制定、审查与修订以电子或非电子形式处理个人数据的政策、过程与程序。⑧促进员工之间的数据保护文化和问责制,并向利益相关者传达个人数据保护政策。⑨确保遵守数据保护法,在个人数据保护政策实施过程中落实监管机构的反馈。⑩直接向董事会报告,并与数据监管机构合作。

为此,DPO 必须具备相关的法律知识与专业技能[55]:①熟悉并了解相关数据保护的法律和实践,特别是对敏感数据的法律保护要求。②熟悉数据控制者或数据处理者处理数据的业务流程和内容,了解其服务机构的业务性质与组织机构。③熟悉数据信息系统和数据安全保护的相关技术,以更好地保护个人数据隐私和安全。④能够在其服务机构内倡导与培育数据保护的组织文化,促使员工遵守数据安全保护法规与流程。

数据控制者或数据处理者可以聘用内部员工或外部机构或个人担任 DPO,但无论是内聘还是外聘 DPO,都要签订数据保护服务合同。同一 DPO 可以在多个机构任职,只要能够胜任工作,且便于联络监管机构、聘任单位和数据主体。此处的同一数据保护官可以是个人,也可以是专门从事数据保护的专业机构。为促使 DPO 更好地履行其职责,DPO 聘用机构或数据控制者、数据处理者需要为 DPO 提供必要的支持,比如:提供履职的资金和基本工作条件,若有需要设立 DPO 团队;保证 DPO 在履职期间不被解雇;要求职能部门给 DPO 履职提供支持;保证 DPO 有充分的时间履行职责;授权 DPO 处理机构数据库或个人数据库中的各项数据;鼓励 DPO 参加各项培训。

建立上述 DPO 制度,不仅可以加强组织内部数据监管,降低数据侵权风险,而且可以增强数据安全治理。虽然 GDPR 要求欧盟成员设置 DPO,但是 DPO 迄今为止还没有写入我国现行法律中。针对科学数据开放共享中的数据安全问题,我国公共机构、研究机构以及大型与中型企业应该借鉴国外经验,对接国际标准,建立 DPO 制度,设置专门的 DPO 岗位,发挥 DPO 在数据安全治理上的重要作用,以便更好地实施科学数据安全治理。

4.2.2 实施数据保护影响评估

数据保护影响评估(data protection impact assessment,DPIA)是一个旨在帮助人们系统地分析、识别项目或计划的数据保护风险并将其最小化的过程,也是一种评估和记录相关数据处理活动、确定活动的风险和减轻或消除这些风险机会的系统方法[56]。DPIA 还可以作为一种工具帮助人们确定最有效的方式来遵守数据保护法定义务和满足人们对隐私保护的期望[57]。DPIA 对象是那些对自然人的权利和自由会产生高风险的数据处理行为,包括数据收集、记录、组织、建构、存储、修改、恢复、查询、披露、传播、分发、使用、清除或销毁等。由于这些数据处理行为往往涉及数据的机密性、完整性和可用性问题,因而,通过实施 DPIA 有助于实现数据安全治理。DPIA 流程包括如下 3 个阶段[58-59]:

(1)准备阶段。该阶段的主要任务是:①考虑是否有必要实施 DPIA。GDPR 要求数据控制者在数据处理给自然人的权利和自由带来高风险时必须实施 DPIA。为了更好地实施科学数据开放共享中的数据安全治理,特别是在存在数据泄露、数据侵权、数据安全隐患时,也应该实施 DPIA。②计划 DPIA,包括定义 DPIA 范围,成立 DPIA 小组。③识别数据处理要求与细节。DPIA 小组或数据控制者必须识别与了解数据处理的目标是什么?数据覆盖哪些学科与地理区域?哪些数据将被收集或处理?是否包括特殊类型数据或敏感信息?如何和从哪里收集这些数据?这些数据将用于何处?将如何处理这些数据?数据格式、标准与适用软件或系统是什么?数据是匿名的还是假名的?数据将如何保存或销毁?数据要保留多长时间?数据共享的方式、范围与对象是什么?用户是否知情同意使用他们的数据?相关的数据安全行业标准、行为准则或公共指南有哪些?突出的数据安全问题有哪些?④识别有关的行动者,包括数据控制者、开发者、组织者、处理者、使用者和其他利益相关者。⑤识别相关法律要求,比如 GDPR、我国《网络安全法》和《科学数据管理办法》对数据安全保护的规制。⑥以标准化程序方式记录准备阶段的结果,包括相关任务和问题。

(2)评估阶段。该阶段的主要任务是:①确定基于数据安全保护目标的评价标准,即按照数据安全保护目标——数据的机密性、完整性和可用性的要求设置评价标准。比如,从机密性要求来看,数据安全保护评价标准必须确保未经授权的数据访问是不允许的;从完整性要求来看,数据安全保护评价标准必须确保要处理的数据是完整无缺的和最新的,是未经修改的、真实的和正确的数据;从可用性要求来看,数据安全保护评价标准必须确保相关数据是可用的、可理解的和可及时处理的,即数据必须可被授权方访问和用适当

的方法进行处理。②识别数据安全风险来源与种类,其中,风险可能来自于数据本身、数据处理过程、数据处理系统与方法等多方面。③确定干预程度与保护水平,设置正常、高和非常高 3 个等级的保护标准。④评估安全风险,即对数据处理行为涉及数据的机密性、完整性和可用性问题,从数据本身、数据处理过程、数据处理系统与方法维度分别进行安全风险评估。⑤确定合适的安全保障措施,比如:采用数据加密技术;限制写权限;比较散列值;定期检查数据完整性;设置最小与最大参考值等。⑥实施已确定的安全保障措施,不过,在实施前需要说明这些措施符合数据安全保护法(如 GDPR)要求。⑦测试和记录评价结果,包括安全保障措施的有效性。⑧制作 DPIA 报告。

(3)复审阶段。在形成 DPIA 报告后,数据监管机构应该评估与审计 DPIA 报告以确保预期的安全保障措施得到实际执行。当数据处理造成的风险发生变化时,需要对 DPIA 进行复审,确保所采用的安全保障措施能够适应这些变化并得到持续监督,使数据安全有保障。

4.2.3 运用数据认证技术

数据认证技术可以用来解决某些数据完整性问题,比如数据失真、数据造假、数据损害、数据篡改、数据丢失等。常用的数据完整性认证技术主要有:基于传统密码学的认证与基于数字水印技术的认证[60]。其中,传统密码学方法主要通过哈希函数产生数字签名,以该签名作为认证信息实现数据判定。数字签名也称电子签名,ISO 7498 - 2 标准将其定义为"附加在数据单元上的一些数据,或是数据单元所做的密码变换,这种数据变换允许数据单元的接收者可以确认数据单元的来源和数据单元的完整性,并保护数据,防止他人伪造"[61]。在科学数据共享活动中,数字签名可进行身份验证以确保已接受到的原始数据的发送者保持不变。它易于运输,不能被其他人复制,并且可以自动加盖时间戳,在消息发送之后,发送者以后也无法轻易修改它。数字签名可用于多种类型的数据传递,无论是否经过加密,都可以使接收者确定发送者的身份并确保数据完整无缺[62]。但该技术也有其不足,当数据发生必要的修改时,则必须抛弃原有的签名并重新计算签名,这将耗费较多的时间。

相对而言,基于数字水印技术的数据认证则具有更强的包容性和抗干扰能力。所谓数字水印技术是指在载体中嵌入一些信息,如作者身份、时间戳、产品属性等,水印的存在形式可以是文字、图形、数列等,当遇到疑似侵权等问题时可以通过算法把水印信息提取出来,从而证明数字产品是否被篡改或者伪造[63]。数据水印技术根据其敏感性可以将其分为脆弱水印和半脆弱水印。脆弱水印非常敏感,主要用于精准认证。在科学数据开放共享中,可以被应用于对十分敏感的多媒体文件的共享,即使共享数据发生了一个比特信息的改变,认证也将无法通过。半脆弱水印则具有更强的适用性,只要内容真实完整,在一定程度上允许常规处理操作,且能把正常的信号处理与恶意篡改区别对待。因此,可以采用半脆弱水印来实施科学数据的版权保护和内容验证,以确保数据的完整性。

4.3 面向数据可用性的治理对策

一项欧洲研究发现,要衡量整个欧洲区域健康不平等在很大程度上取决于区域一级可靠和可比数据的可用性;消除"数据差距"是消除欧盟国家之间和欧盟国家内部"健康差距"的条件[64]。科学数据的可用性对于其他行业如航空工业[65]、制药工业[66]等的发展同样至关重要。在科学数据开放共享中,科学数据管理政策、科学数据质量、开放共享平台等都能影响科学数据的可用性。因此,可以采取制定科学数据可用性政策或发布数据可用性声明、提高科学数据质量、建立科学数据统一开放共享平台等多项措施来增强数据可用性,以实现数据安全治理的目的。

4.3.1 制定科学数据可用性政策或发布数据可用性声明

在开放数据运动中,许多政府机构、研究机构和出版社制定了本机构数据可用性政策或发布数据可用性声明,以促进科学数据的开放共享与利用。2019 年 12 月 23 日,美国管理和预算办公室(Office of Management and Budget, OMB)发布了《联邦数据战略与 2020 年行动计划》。"将数据作为战略资产加以利用"成为该战略的核心目标。为此,该战略要求按照承担责任(如实施有效的数据管理和治理、采用可靠的数据安全措施、保护个人私隐、保持承诺的机密性、确保适当的访问与使用)、促进透明度、确保相关性(如保护数据的质量和完整性,确认数据是适当的、准确的、客观的、可获得的、有用的、可理解的和及时的)等 10 项原则,开展 3 类 40 项数据管理实践,包括:①建立重视数据和促进数据公共使用的文化,如支持数据使用、使用数据来指导决策、准备共享、跨机构连接数据功能等;②控制、管理与保护数据,如保护数据完整性、传递数据的真实性、为数据资产开列清单、确认数据资产的价值、维护数据文档、利用数据标准、与数据管理需求保持一致、

盛小平，郭道胜. 科学数据开放共享中的数据安全治理研究[J]. 图书情报工作,2020,64(22):25－36.

加强数据保护以及在州、地方、部落政府和联邦机构之间共享数据等；③促进有效和适当的数据使用，如促进广泛的访问、审查数据发布是否存在披露风险等[67]。由此看来，该战略不仅从国家层面提供了如何管理与使用联邦政府数据的指南，而且为联邦政府数据的可用性提供了保障，有助于提升联邦政府数据的可用性，促进数据共享。

发布数据可用性声明往往成为许多出版社保障数据可用性的首要选择。一些研究资助者，如英国研究理事会等，要求在出版物中包含数据可用性声明。《自然》鼓励通过不同方式来提供数据可用性声明，比如：在当前研究中生成和/或分析的数据集可以在《自然》指定的存储库中获得；在当前研究中产生和/或分析的数据集可在合理请求下从通讯作者处获得；本研究中产生或分析的所有数据均包含在本文及其补充文档中；由于存在数据不能公开的原因，通讯作者可以在合理请求下提供当前研究中产生和/或分析的数据集[68]。其他世界顶级期刊往往也提供数据可用性声明来助力开放科学研究和确保科学数据的可发现、可验证和重用[69]。总之，制定科学数据可用性政策或发布数据可用性声明有助于增强科学数据的可用性，从而改善科学数据的安全治理。

4.3.2 提高科学数据质量

已有研究证实，数据质量与数据可用性之间存在正相关关系[70]。也就是说，通过提高数据质量可以有效增强数据的可用性。一方面，反映数据质量的属性有多种多样，比如准确性、机密性、完整性、可用性、一致性、及时性、关联性、有效性等。可用性虽然仅是数据质量的一种属性，但是它与其他数据质量属性都有内在联系。因此，增强科学数据的可用性可以从提高数据的多方面属性入手。另一方面，数据质量与数据的生产、收集、组织、存储、发布（或出版）都有紧密联系。共享是数据生命周期中的重要一环，本身包含收集、组织、发布、传播和利用等过程[71]。科学数据开放共享主要包括科学数据的开放发布、开放获取、开放存储、开放利用。因此，在科学数据开放共享过程中，提高科学数据质量需要关注开放共享过程。此外，科学数据是人们在各项科学研究、生产与管理实践中产生的，与数据的生产者、组织者、发布者、传播者、管理者、利用者都有直接关系。因此，提高科学数据质量需要考虑利益相关者的全员参与。总而言之，迫切需要实施全面数据质量管理（total data quality management，TDQM）来增强数据的可用性。

TDQM 是运用全面质量管理思想对数据或数据产品进行有效管理以提高其质量与效用的一种管理方法。结合现有观点[72－73]，可以构建 TDQM 流程（见表 4）以提高数据质量。在 TDQM 实施过程中，至关重要的是：①定义科学数据质量需求，特别是数据的准确性、机密性、完整性、可用性、一致性、及时性、关联性、有效性等，以便根据这些质量需求实施全面质量管理。②明确 TDQM 的利益相关者，比如数据的生产者、供应者、组织者、传播者、管理者与使用者，以便实现利益相关者的全员管理。③逐一执行 TDQM 流程，实施螺旋递进式的 TDQM 循环，从而不断提高包括可用性在内的数据质量。④明确数据质量管理中的数据治理任务，为数据质量建立数据治理框架，从而确保 TDQM 的顺利实施。

表 4 TDQM 流程

阶段	流程实现步骤
定义阶段	步骤 1：建立 TDQM 团队，确定利益相关者 步骤 2：确定数据产品特征与数据产品属性特征 步骤 3：确定数据质量需求，包括确定重要的数据质量维度 步骤 4：确定数据质量管理中的数据治理任务
度量阶段	步骤 5：确定数据质量测量指标 步骤 6：度量和呈现数据质量，以帕累托图显示结果
分析阶段	步骤 7：描述具体问题：根据测量结果描述具体的数据质量问题 步骤 8：分析问题：与整个质量团队一起绘制原因图来分析问题
改进阶段	步骤 9：形成解决方案：针对质量问题思考可能的解决方案 步骤 10：选择解决方案：根据事先确定的度量标准对备选方案进行评分，并选择 步骤 11：实施数据质量改进行动计划：为团队成员分配特定的任务来实施解决方案 步骤 12：持续改进：检查行动计划进展，持续改进数据质量，开展新一轮的 TDQM

4.3.3 构建基于数据联盟的国家科学数据中心

科学数据开放共享平台在提供可利用的科学数据方面具有无可替代的关键作用。我国于 2017 年已经建成包括“国家人口与健康科学数据共享服务平台”等在内的 8 个国家科学数据共享平台，又于 2019 年在原有科学数据类国家平台的基础上提出了建设包括“国家高能物理科学数据中心”等在内的 20 个国家科学数据中心，并把其作为优化调整国家科技资源共享服务平台、完善科技资源共享服务体系、推动科技资源向社会开放共享的战略选择。不过，在我国科学数据共享平台建设过程中，一些平台存在数据可获取性与引用率较低的问题[74]。如今的国家高能物理科学数据中心、国家基因组科学数据中心等国家科学数据中心虽然在特定的学科领域可以为注册人员提供科学数据的访问，但是并没有实现科学数据的开放共享。更何况这些国家科学数据中心彼此之间没有建立关联，

更没有集成起来形成统一的"国家科学数据中心"。因此,这些中心的科学数据仍存在可用性不高的问题。

不同于国内科学数据中心,由莫纳什大学牵头,联合澳大利亚国立大学、联邦科学与技术研究组织(Commonwealth Scientific and Industrial Research Organisation, CSIRO)组建的澳大利亚国家数据服务中心(Australian National Data Service, ANDS)不仅负责管理澳大利亚的科学数据,而且通过下属的"澳大利亚研究数据"(Research Data Australia)门户开放共享来自100多个澳大利亚研究机构、政府机构和大学的研究数据,涵盖自然科学、社会科学、艺术和人文学科等多学科领域[75]。它是一种基于数据联盟的运营模式,在提高科学数据的可用性与开放性方面树立了成功的典范。这种模式实质上是国家科学数据中心(或国家科学数据共享平台)联合不同科学数据的生产者、提供者、组织者、管理者组成科学数据联盟,共同参与科学数据的共享与利用活动[76]。用户通过国家科学数据中心(或国家科学数据共享平台)不仅可以访问该中心拥有的各种科学数据,而且可以利用国家科学数据共享平台建立的联盟成员数据集索引,访问存储在异地联盟成员机构知识库的科学数据,从而极大促进科学数据的开放共享。因此,通过构建基于数据联盟的国家科学数据中心,有助于解决目前我国科学数据可用性水平不高的问题,改善科学数据的安全治理。

5 结语

保障数据安全是实施科学数据开放共享不可回避的关键问题。科学数据开放共享中的数据安全问题集中体现在数据机密性问题、完整性问题和可用性问题3个方面,亟需从法律、政策、制度、管理、技术与平台等维度采取多种数据安全治理措施来处理这些问题,从而构建科学数据开放共享的数据安全治理体系。不过,本文提出的观点仍停留在理论探讨上,还需在实践中进一步验证与完善,由此实现有效治理我国科学数据开放共享中的数据安全问题和提高我国数据治理水平与国家治理能力的目的。

参考文献:

[1] 石英村. 全球数据安全治理态势与产业趋势分析[J]. 信息安全与通信保密,2019, 41(4):35-37.

[2] 张汉青. 大数据时代数据安全需要多级保护[N]. 经济参考报,2019-05-09(7).

[3] 盛小平,武彤. 国内外科学数据开放共享研究综述[J]. 图书情报工作,2019,63(17): 6-14.

[4] PEREIRA S, GIBBS R A, MCGUIRE A L. Open access data sharing in genomic research [J]. Genes, 2014,5(3):739-747.

[5] WIEBE A, DIETRICH N. Open data protection: study on legal barriers to open data sharing -data protection and PSI [M]. Göttingen: Universitätsverlag Göttingen, 2017.

[6] LITTLE D D B, FARMER S, EL-HILALI OU. Digital data integrity: the evolution from passive protection to active management [M]. West Sussex: John Wiley & Sons Ltd, 2007.

[7] 李善青,郑彦宁,邢晓昭,等. 科学数据共享的安全管理问题研究[J]. 中国科技资源导刊,2019, 51(3):11-17.

[8] 杜跃进. 数据安全治理的几个基本问题[J]. 大数据,2018,4(6):85-91.

[9] 王世晞,张亮,李娇娇. 大数据时代下的数据安全防护——以数据安全治理为中心[J]. 信息安全与通信保密,2020,42(2):82-88.

[10] HILL D G. Data protection: governance, risk management, and compliance[M]. Boca Raton: CRC Press,2010.

[11] 付霞,付才. 新时代数据安全风险的法律治理[J]. 长江大学学报(社会科学版), 2019,42(2): 58-61.

[12] LIVRAGA G,TORRA V, ALDINI A, et al. Data privacy management and security assurance [M]. Cham: Springer International Publishing AG, 2016.

[13] TAMANE S, SOLANKI V K, DEY N. Privacy and security policies in big data[M]. Hershey: IGI Global, 2017.

[14] MOSLEY M, BRACKETT M, EARLEY S, et al. The DAMA guide to the data management body of knowledge (DAMA-DM-BOK)[M]. Bradley Beach: Technics Publications, 2009:151.

[15] SOLMS S H,SOLMS R. Information security governance[M]. New York: Springer, 2009:24.

[16] 陈磊. 拨开云雾见天日——数据安全治理体系[J]. 安全月刊,2019(10):4-10.

[17] European Commission. Guidelines on open access to scientific publications and research data in Horizon 2020, Version 3.2[EB/OL]. [2020-06-06]. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

[18] DENNING D E R. Cryptography and data security [M]. Massachusetts: Addison-Wesley Publishing Company, 1982: V, 7.

[19] MORAN R, LEVINGER J. Oracle security overview 10g release 1 (10.1) [R/OL]. [2020-06-06]. https://docs.oracle.com/cd/B12037_01/network.101/b10777.pdf.

[20] 冯登国. 大数据安全与隐私保护[M]. 北京:清华大学出版社,2018:5.

[21] CALDER A, WATKINS S. IT governance: an international guide to data security and ISO27001/ISO27002 [M]. 6th ed. London: Kogan Page Limited, 2015:10.

[22] National Academies of Sciences, Engineering, and Medicine. Open science by design: realizing a vision for 21st century research [M]. Washington, DC: The National Academies Press,2018:50-51

[23] 叶润国,陈雪秀. 政府数据开放共享安全保障问题与建议[J]. 信息技术与标准化,2016,58(6): 22-25,34.

[24] PENG C,SONG X,JIANG H, et al. Towards a paradigm for open and free sharing of scientific data on global change science in China

[J/OL]. Ecosystem health and sustainability, 2016, 2(5): e01225. [2020-06-06]. https://esajournals.onlinelibrary.wiley.com/doi/epdf/10.1002/ehs2.1225.
[25] 刘润达,孙九林,廖顺宝. 科学数据共享中数据授权问题初探[J]. 情报杂志,2010,29(12):15-18.
[26] STAGARS M. Open data in Southeast Asia[M]. Singapore: Palgrave Macmillan, 2016:17-20.
[27] JANSSEN M, CHARALABIDIS Y, ZUIDERWIJK A. Benefits, adoption barriers and myths of open data and open government[J]. Information systems management, 2012, 29(4):258-268.
[28] 温亮明,张丽丽,黎建辉. 大数据时代科学数据共享伦理问题研究[J]. 情报资料工作,2019,40(2):38-44.
[29] 张一鸣. 数据治理过程浅析[J]. 中国信息界,2012,10(9):15-17.
[30] Committee on Science, Engineering, and Public Policy(U.S.), Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age. Ensuring the integrity, accessibility, and stewardship of research data in the digital age[M]. Washington, DC: The National Academies Press, 2009:96.
[31] KUULA A, BORG S. Open access to and reuse of research data - the state of the art in Finland[M]. Tampere: Finnish Social Science Data Archive,2008:11-12.
[32] 辛一. 九省份科学数据共享平台网站建设比较研究[J]. 中国科技资源导刊,2019,51(3):18-23.
[33] SEDRANSK N,YOUNG L J,KELNER K L, et al. Make research data public? Not always so simple: a dialogue for statisticians and science editors[J]. Statistical science, 2010,25(1):41-50.
[34] POTER M E. Competitive advantage: creating and sustaining superior performance[M]. New York: The Free Press, 1985:36-43.
[35] 中华人民共和国国家安全法(全文)[EB/OL]. [2020-06-06]. http://news.sina.com.cn/c/2015-07-01/220132055212.shtml.
[36] 中华人民共和国网络安全法[EB/OL]. [2020-06-06]. http://www.xinhuanet.com//zgjx/2016-11/08/c_135813275.htm.
[37] 周羽. 全国人大代表魏明:加快制定《数据安全法》[EB/OL]. [2020-06-06]. https://www.sohu.com/a/397151630_362042.
[38] 国务院办公厅. 国务院办公厅关于印发科学数据管理办法的通知[EB/OL]. [2020-06-06]. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.
[39] 国家互联网信息办公室. 国家互联网信息办公室关于《数据安全管理办法(征求意见稿)》公开征求意见的通知[EB/OL]. [2020-06-06]. http://www.gov.cn/xinwen/2019-05/28/content_5395524.htm.
[40] Data classification guide[EB/OL]. [2020-06-06]. https://www.spirion.com/data-classification/.
[41] 国务院. 国务院关于印发促进大数据发展行动纲要的通知[EB/OL]. [2020-06-06]. http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm? url_type=39&object_type=webpage&pos=1.
[42] 工业和信息化部办公厅. 工业和信息化部办公厅关于印发《工业数据分类分级指南(试行)》的通知[EB/OL]. [2020-06-06]. http://www.miit.gov.cn/n1146295/n1652858/n1652930/n3757016/c7772152/content.html.
[43] UNSW. Research data governance & materials handling policy [EB/OL]. [2020-06-06]. https://www.gs.unsw.edu.au/policy/documents/researchdatagovernancepolicy.pdf.
[44] AWS. Data classification: secure cloud adoption[EB/OL]. [2020-06-06]. https://d1.awsstatic.com/whitepapers/compliance/AWS_Data_Classification.pdf.
[45] Berkeley Information Security Office. How to classify research data [EB/OL]. [2020-06-06]. https://security.berkeley.edu/education-awareness/best-practices-how-tos/how-classify-research-data.
[46] UNSW. Data classification standard[EB/OL]. [2020-06-06]. https://www.gs.unsw.edu.au/policy/documents/datastandard.pdf.
[47] SWEENEY L, CROSAS M, BAR-SINAI M. Sharing sensitive data with confidence: the datatags system[EB/OL]. [2020-06-06]. https://techscience.org/a/2015101601/download.pdf.
[48] 付钰,俞艺涵,吴晓平. 大数据环境下差分隐私保护技术及应用[J]. 通信学报, 2019,40(10): 157-168.
[49] The Royal Society. Israel-UK privacy and technology workshop note of discussions [EB/OL]. [2020-06-06]. https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/israel-uk-privacy-and-technology-workshop-note.pdf? la=en-GB&hash=218915A3D5AA244D333A22D104882551.
[50] ALAMEDA T. What are PET technologies?: how to maximize data value while preserving privacy [EB/OL]. [2020-06-01]. https://www.bbva.com/en/what-are-pet-technologies-how-to-maximize-data-value-while-preserving-privacy/.
[51] INPHER. What is secure multiparty computation? [EB/OL]. [2020-06-02]. https://www.inpher.io/technology/what-is-secure-multiparty-computation.
[52] DOBRAN B. What is data integrity? why your business needs to maintain it[EB/OL]. [2020-06-03]. https://phoenixnap.com/blog/what-data-integrity.
[53] MONTEZUMA L A. Why should a data protection officer be global? [EB/OL]. [2020-06-06]. https://iapp.org/news/a/why-should-a-data-protection-officer-be-global/.
[54] LAMBERT P. The data protection officer: profession, rules, and role [M]. Boca Raton: CRC Press, 2017:45-46.
[55] 刘江山. 欧盟通用数据保护条例中的数据保护官制度[J]. 中国科技论坛,2019(12):173-179.
[56] Freeprivacypolicy. GDPR data protection impact assessments[EB/OL]. [2020-06-06]. https://www.freeprivacypolicy.com/blog/gdpr-data-protection-impact-assessment/.
[57] Central London Community Healthcare. Data protection impact assessment [EB/OL]. [2020-06-06]. https://clch.nhs.uk/about-us/publications/data-protection-impact-assessment-dpia-summaries.
[58] BIEKER F, FRIEDEWALD M, HANSEN M, et al. A process for data protection impact assessment under the Eropean General Data Protection Regulation [J]. Lecture notes in computer science, 2016,9857:21-37.

[59] BIEKER F, MARTIN N, FRIEDEWALD M, et al. Data protection impact assessment: a hands-on tour of the GDPR's most practical tool[C]//HANSEN M, KOSTA E, NAI-FOVINO I, et al. Privacy and identity management: the smart revolution. Cham: Springer International Publishing AG, 2018:207 - 220.

[60] 张鑫港,闫浩文,张黎明. 一种用于 DEM 数据认证与篡改定位的感知哈希算法[J]. 地球信息科学学报,2020,22(3):379 - 388.

[61] 李拴保. 信息安全基础[M]. 北京:清华大学出版社,2014.

[62] SHOEB Z H, SOBBAN M A. Authentication and authorization: security issues for institutional digital repositories[J]. Library philosophy and practice, 2010, 12(5):1 - 6.

[63] 谭慧. 数字水印技术及其应用[J]. 信息与电脑(理论版), 2018,12(13):221 - 222,225.

[64] COSTA C, FREITAS A, STEFANIK I, et al. Evaluation of data availability on population health indicators at the regional level across the European Union[EB/OL]. [2020 - 06 - 04]. https://pophealthmetrics. biomedcentral. com/articles/10. 1186/s12963-019-0188-6.

[65] SHEHAB E, BOUIN-PORTET M, HOLE R, et al. Enhancing digital design data availability in the aerospace industry[J]. CIRP journal of manufacturing science and technology, 2010,2(4): 240 - 246.

[66] HOPKINS A M, ROWLAND A, SORICH M J. Data sharing from pharmaceutical industry sponsored clinical studies: audit of data availability[EB/OL]. [2020 - 06 - 04]. https://bmcmedicine. biomedcentral. com/articles/10. 1186/s12916-018-1154-z.

[67] The President's Management Agenda Team. Federal data strategy 2020 action plan[EB/OL]. [2020 - 06 - 04]. https://strategy. data. gov/assets/docs/2020-federal-data-strategy-action-plan. pdf.

[68] Springer Nature. Data availability statements[EB/OL]. [2020 - 06 - 04]. https://www. springernature. com/gp/authors/research-data-policy/data-availability-statements/12330880.

[69] Editorial. On data availability, reproducibility and reuse [J]. Nature cell biology, 2017, 19(4):259 - 259.

[70] 丁小欧,王宏志,张笑影, 等. 数据质量多种性质的关联关系研究[J]. 软件学报, 2016,27(7): 1626 - 1644.

[71] 毕达天,曹冉,杜小民. 科学数据共享研究现状与展望[J]. 图书情报工作,2019, 63(24):69 - 77.

[72] WIJNHOVEN F, BOELENS R, MIDDEL R, et al. Total data quality management: a study of bridging rigor and relevance[EB/OL]. [2020 - 06 - 05]. https://ris. utwente. nl/ws/portalfiles/portal/47275011/Wijnhoven07total. pdf.

[73] BOELENS R. A product-attribute approach to total data quality management[EB/OL]. [2020 - 06 - 06]. http://essay. utwente. nl/57694/1/scriptie_Boelens. pdf.

[74] 司莉,华小琴. 我国科学数据共享平台的服务效能分析[J]. 图书馆工作与研究, 2014, 36(4):24 - 26.

[75] Australian National Data Service. Research data Australia[EB/OL]. [2020 - 06 - 08]. https://www. ands. org. au/online-services/research-data-australia.

[76] 刘润达,赵辉,李大玲. 科学数据共享平台之数据联盟模式初探[J]. 中国基础科学,2010,12(6): 27 - 32.

作者贡献说明:
盛小平:撰写与修订论文;
郭道胜:参与论文初稿撰写。

Research on Data Security Governance in Open Sharing of Scientific Data

Sheng Xiaoping Guo Daosheng
School of Library, Information and Archives, Shanghai University, Shanghai 200444

Abstract: [**Purpose/significance**] This paper reveals the data security problems in the open sharing of scientific data, and puts forward corresponding governance countermeasures, so as to promote the practice of open sharing of scientific data in China better. [**Method/process**] By means of normative analysis, this paper analyzed and defined the data security problems in the open sharing of scientific data, and then discussed the governance measures for the security of scientific data from the three dimensions of confidentiality, integrity and availability. [**Result/conclusion**] There are a lot of security problems in data confidentiality, integrity and availability in the open sharing of scientific data. The problems of data confidentiality can be governed by three measures including strengthening data security legislation, establishing scientific data classification standards and systems, and making full use of privacy enhancing technologies. The problems of data integrity can be governed by three measures including establishing a data protection officer system, implementing data protection impact assessment, and using data authentication technologies. The problems of data availability can be governed by three measures including formulating policies on the availability of scientific data, improving the quality of scientific data, and building a national scientific data center based on data alliance.

Keywords: scientific data open sharing data security security governance